On the biosynthetic origin of carminic acid

Rasmussen, Silas A.; Kongstad, Kenneth T; Khorsand-Jamal, Paiman; Kannangara, Rubini Maya; Nafisi, Majse; Van Dam, Alex; Bennedsen, Mads; Madsen, Bjørn; Okkels, Finn; Gotfredsen, Charlotte H.

Total number of authors:
15

Published in:
Insect Biochemistry and Molecular Biology

Link to article, DOI:
10.1016/j.ibmb.2018.03.002

Publication date:
2018

Document Version
Peer reviewed version

Accepted Manuscript

On the biosynthetic origin of carminic acid

Silas A. Rasmussen, Kenneth T. Kongstad, Paiman Khorsand-Jamal, Rubini Maya Kannangara, Majse Nafisi, Alex Van Dam, Mads Bennedsen, Bjørn Madsen, Finn Okkels, Charlotte H. Gotfredsen, Dan Staerk, Ulf Thrane, Uffe H. Mortensen, Thomas O. Larsen, Rasmus J.N. Frandsen

PII: S0965-1748(18)30111-5

DOI: 10.1016/j.ibmb.2018.03.002

Reference: IB 3039

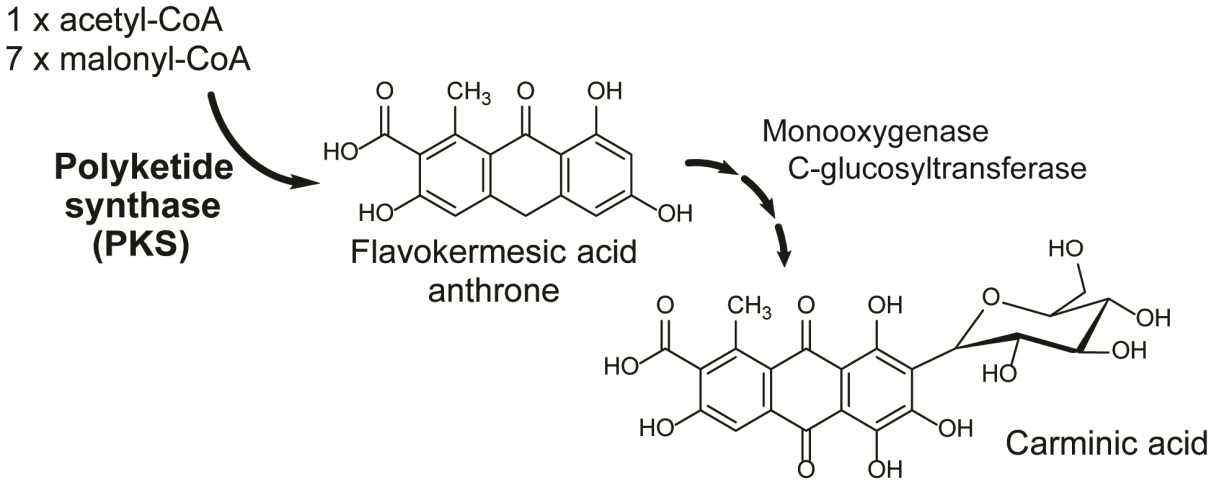
To appear in: *Insect Biochemistry and Molecular Biology*

Received Date: 8 July 2017

Revised Date: 9 March 2018

Accepted Date: 13 March 2018

Please cite this article as: Rasmussen, S.A., Kongstad, K.T., Khorsand-Jamal, P., Kannangara, R.M., Nafisi, M., Van Dam, A., Bennedsen, M., Madsen, Bjø., Okkels, F., Gotfredsen, C.H., Staerk, D., Thrane, U., Mortensen, U.H., Larsen, T.O., Frandsen, R.J.N., On the biosynthetic origin of carminic acid, *Insect Biochemistry and Molecular Biology* (2018), doi: 10.1016/j.ibmb.2018.03.002.



On the biosynthetic origin of carminic acid

Silas A. Rasmussen [a], Kenneth T. Kongstad [b], Paiman Khorsand-Jamal [b,c], Rubini Maya Kannangara [c,d], Majse Nafisi [c,d], Alex Van Dam [a,e], Mads Bennedsen [c], Bjørn Madsen [c], Finn Okkels [c,f], Charlotte H. Gotfredsen [g], Dan Staerk [b], Ulf Thrane [a], Uffe H. Mortensen [a], Thomas O. Larsen [a,*] and Rasmus J.N. Frandsen [a,*]

[a] *Department of Bioengineering and Biomedicine, Technical University of Denmark, Søltofts Plads, Building 221 and 223, 2800 Kongens Lyngby, Denmark.*

[b] *Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen, Denmark.*

[c] *Chr. Hansen A/S, Boege Alle 10-12, 2970 Hoersholm, Denmark.*

[d] *Plant Biochemistry Laboratory, Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark.*

[e] *Department of Biology, University of Puerto Rico, PO Box 9000, Mayagüez, PR 00681-9000, Puerto Rico.*

[f] *Current adresse: Actabio ApS, Kongemarken 11, 4000 Roskilde, Denmark*

[g] *Department of Chemistry, Technical University of Denmark, Kemitorvet, Building 207, 2800 Kongens Lyngby, Denmark.*

* Corresponding author: Phone: +45 45252632, E-mail: tol@bio.dtu.dk and Phone: +45 45252708, E-mail: rasf@bio.dtu.dk

ABSTRACT

The chemical composition of the scale insect *Dactylopius coccus* was analyzed with the aim to discover new possible intermediates in the biosynthesis of carminic acid. UPLC-DAD/HRMS analyses of fresh and dried insects resulted in the identification of three novel carminic acid analogues and the verification of several previously described intermediates. Structural elucidation revealed that the three novel compounds were desoxyerythrolaccin-*O*-glucosyl (**DE-*O*-Glc*p***), 5,6-didehydroxyerythrolaccin 3-*O*-β-D-glucopyranoside (**DDE-3-*O*-Glc*p***), and flavokermesic acid anthrone (**FKA**). The finding of **FKA** in *D. coccus* provides solid evidence of a polyketide, rather than a shikimate, origin of coccid pigments. Based on the newly identified compounds, we present a detailed biosynthetic scheme that accounts for the formation of carminic acid (**CA**) in *D. coccus* and all described coccid pigments which share a flavokermesic acid (**FK**) core. Detection of coccid pigment intermediates in members of the *Planococcus* (mealybugs) and *Pseudaulacaspis* genera shows that the ability to form these pigments is taxonomically more widely spread than previously documented. The shared core-**FK**-biosynthetic pathway and wider taxonomic distribution suggests a common evolutionary origin for the trait in all coccid dye producing insect species.

Keywords: Carminic acid, carmine, anthraquinones, coccid pigment, polyketide, insects, *Dactylopius coccus*, biosynthesis

1. Introduction

Pigments derived from insects and especially coccids (scale insects) have been used by humans since ancient times for dyeing textiles, in cosmetics and in paints, and for coloring foods (Donkin, 1977). The most commonly used coccid dyes include kermesic acid (**KA**), laccaic acids (**LA**) and carminic acid (**CA**), which share a red color hue due to a similar chromophore structure (Łagowska and Golan, 2009). The compounds, or combinations of these, have been reported to be produced by several species of distantly related scale insects (*Hemiptera: Coccoidea*). Mainly five species, namely *Porphyrophora hamelii* (Armenian/Ararat cochineal), *Kermes vermilio* (kermes), *Porphyrophora polonica* (Polish cochineal), *Dactylopius coccus* (Mexican cochineal) and *Kerria lacca* (Indian lac insect) have at various points in history, and at different geographical localities, been utilized by humans for large scale production of coccid dyes (Donkin, 1977). Carminic acid and its aluminum salt carmine (E120) is by many considered as the pinnacle of coccid dyes, based on its hue, light, temperature, and oxidation stability, and the yields by which it can be obtained from natural sources (Dapson, 2007). **CA** is known to be produced by *P. hamelii* (Asia Minor), *P. polonica* (Europe), and *D. coccus* (Meso and South America), all of which have served as sources for the compound (Wouters and Verhecken, 1989). Present day production is based on *D. coccus* due to its exceptional high pigment content (16-22% of dry weight), low fat content, and the ease by which the insect can be cultured and harvested from leaves of *Opuntia* cacti (Donkin, 1977; Downham and Collins, 2000). A thorough introduction to the historical use and geopolitical role of carmine is given by Dapson (Dapson, 2007).

Although insect-derived pigments have been utilized by humans for millennia and remain of significant value within the food colorant market, the underlying biochemistry for their production remains largely unknown. The coccid dyes, such as **CA**, have by many authors been categorized as polyketides solely based on their structure (Morgan, 2010; Cameron et al., 1978; Pankewitz and Hilker, 2008; Brown, 1975). The biosynthesis mechanisms of formation of polyketides, via the successive condensation of acetyl-CoA and malonyl-CoA units catalyzed by polyketide synthases (PKSs), is well described in bacteria, fungi, and plants (Staunton and Weissman, 2001). However, no animal PKSs have yet been biochemically characterized, even though many insect species are known to contain compounds that potentially may be synthesized via the polyketide pathway. In most cases, the putative polyketides contents have been ascribed to the sequestering of precursors, or the finished compounds, from the insects' diet (Pankewitz and Hilker, 2008). This situation is seen in *Timarcha* spp. (leaf beetles) which accumulates anthraquinones from its host plant *Galium* spp. (Rubiaceae) (Petitpierre, 1981), and in *Laetilia coccidivora* (pyralid moth) and *Hyperaspis trifurcate* (coccinellid beetle) larva that accumulate **CA** by predating on *Dactylopius* spp. (Eisner et al., 1994). In other cases, the origins of the detected polyketides in insects have been linked to the activity of

endosymbiotic bacteria, such as the production of pederin, a polyketide-peptide hybrid, in *Paederus* spp. (rove beetles), which depends on an endosymbiotic bacterium related to *Pseudomonas aeruginosa* (Piel, 2002; Kellner, 2002). A second example is the facultative endosymbiotic *Rickettsiella* spp. responsible for the production of the polyketide viridaphin A1 in various aphids (*Acyrthosiphon pisum* and *Megoura crassicauda*) (Tsuchida et al., 2010; Horikawa et al., 2011). The biosynthetic origin of coccid pigments in scale insects, however, remains a mystery. The hypothesis that coccid dyes are polyketides has solely been based on their chemical structure, even though the polyketide class is characterized by a shared mode of synthesis rather than shared structural features (Staunton and Weissman, 2001). In fact, for the anthraquinone core of **CA**, one could envision that this is formed via the shikimate based chorismate/*O*-succinyl benzoic acid pathway, as described for lucidi, alizarin, and morindone in rubiaceous plants (Leistner, 1973; Han et al., 2001). If CA is formed by this pathway it would not qualify as a polyketide and its formation would not dependent on a PKS. However, it is possible to distinguish between the two alternative pathways as the polyketide-based synthesis would include a unique anthrone intermediate, which is not found in the shikimate-based pathway, where the anthraquinone is formed directly.

The present study aims at increasing our understanding of the **CA** origin and its biosynthesis in *D. coccus*. We also aim to elaborate on the previously proposed links to other biosynthetic pathways responsible for the production of coccid dyes within the *Coccoidea* superfamily. In our study, we report the presence of flavokermesic acid anthrone (**FKA**) in *D. coccus*, which strengthens the hypothesis that coccid pigments are formed via the polyketide pathway, as **FKA** is the first cyclic intermediate in a polyketide-based pathway.

2. Materials and methods

2.1. Biological material

Adult *D. coccus* specimens were collected from *Opuntia* cacti pads on the Canary Islands, Lanzarote, near the village of Guatiza, June 2012. The insects were transported to Denmark either as live specimens on cacti pads or as dead specimens stored on dry ice. Additional *D. coccus* insects were collected from cacti pads near the city of Arequipa Peru, August 2012, flash frozen in liquid nitrogen, and shipped to Denmark on dry ice. Commercially available dried *D. coccus* insects were supplied by Chr. Hansen A/S.

Coccus hesperidium, Pseudococcus longispinus, Palmicultor browni, and *Pseudaulacaspis pentagona* were collected in the greenhouses of the Botanical Garden (Natural History Museum of Denmark, University of Copenhagen) in Copenhagen in June 2014, and identified using the latest available identification keys (Dooley and Dones, 2015; Miller et al., 2014).

2.2. Instrumentation

Chemical analysis of *D. coccus* samples was performed using three different LC-MS setups. UPHLC-DAD-HRMS was performed on a maXis G3 QTOF mass spectrometer (Bruker Daltronics, Bremen, Germany) equipped with an electrospray ionization source coupled to an Ultima 3000 UHPLC-DAD (Dionex). Separation was performed on a Kinetex C_{18} column (150 × 2.1 mm, 2.6 µm, Phenomenex Inc., Torrance, CA, USA) maintained at 40 °C using a linear H_2O-acetonitrile gradient consisting of A: milliQ H_2O containing 10 mM formic acid and B: acetonitrile containing 10 mM formic acid from 10 to 100% B in 10 min with a flow rate of 400 µL min^{-1}. The **FK** anthrone was detected on a HPLC-DAD-HRMS system consisting of an Agilent 1200 chromatograph comprising quaternary pump, degasser, thermostatted column compartment, autosampler, and photodiode array detector (Agilent Technology, Santa Clara, CA, USA) and a Bruker micrOTOF-Q II mass spectrometer (Bruker Daltonik, Bremen, Germany) equipped with an electrospray ionization source and operated via a 1:99 flow splitter. Analyses were performed at 40 °C on a Luna $C_{18}(2)$ reversed-phase column (150 × 4.6 mm, 3 µm particle size, 100 Å pore size, Phenomenex Inc., Torrance, CA, USA) with a flow rate of 800 µL min^{-1}. HPLC solvent A consisted of H_2O-acetonitrile 95:5 (v/v) with 0.1% formic acid and solvent B consisted of acetonitrile-H_2O 95:5 (v/v) with 0.1 % formic acid. Separation was obtained using a linear gradient from 0 to 100% B in 20 minutes. Mass spectra were acquired in negative ionization mode. The search for coccid dye intermediates in the different scale insect species was performed on a 6540 Ultra High Definition UHD Accurate Mass Quadrupole Q-TOF LC/MS system (Agilent Technology, Santa Clara, CA, USA). Separation of the analytes was conducted on a Kinetex XB-C_{18} (100 x 4.6 mm i.d. 2.6 µm, Phenomenex Inc., Torrance, CA, USA), column maintained at 35 °C. The analytes were eluted with a flow rate of 400 400 µL min^{-1} using a water-acetonitrile gradient consisting of the following steps: 100% water for 50 s followed by an gradual increase to 18.6% over 60 s, to 37.8% over 60 s, to 52.2% over 120 s, to 54.2% over 70 s, to 90% over 120 s, to 100% over 120 s followed by 60 s at 100% acetonitrile. The column was reconstituted with 100% water for 110 s prior to injection of the subsequent sample. NMR spectra of 5,6-didehydroxyerythrolaccin 3-*O*-β-D-glucopyranoside (**DDE-3-*O*-Glc*p***), desoxyerythrolaccin *O*-glucopyranoside (**DE-*O*-Glc*p***), and **dcII** were recorded on a Varian Inova 500 MHz (Varian Inc., Palo Alto, California) using a 5-mm probe. Samples were dissolved in 500 µl DMSO-d_6 and referenced to δ_H at 2.50 ppm and δ_C at 39.5 ppm. The NMR spectrum of flavokermesic acid anthrone (**FKA**) was recorded on a Bruker Avance III HD 600 MHz NMR spectrometer (^{1}H operating frequency 600.13 MHz) equipped with a cryogenically cooled 5-mm CPDCH probe-head (Bruker Biospin, Rheinstetten, Germany). The sample was dissolved in acetone-d_6 and referenced to δ_H 2.05 ppm and δ_C 29.84 ppm. Following structural elucidation of the described compounds, their presence in the original samples was verified using targeted MS analysis.

Chiral GC-MS was performed using 10 µg **DDE-3-*O*-Glc** that was hydrolyzed in 10% aqueous HCl for 90 min at 90 °C, dried by a steam of N_2 and dissolved in 40 µL dry pyridine followed by 10 µL N-methyl-bis-trifluoroacetamide (MBTFA) (GC-grade, 99%, Sigma-Aldrich) and heating to 65 °C for 40 min. The sample was cooled to room temperature and subsequently analyzed on a CP-ChiraSil-L-Val GC column (25 m × 0.25 mm × 0.12 µm, Agilent Technology, Santa Clara, CA, USA) programmed to 70−150 °C at 4 °C min^{-1} using an HP 6890 series GC system and Agilent 5973 mass selective detector. The sample was compared to the standards of D- and L-glucose (Sigma-Aldrich, St. Louis, MO, USA).

2.3. Extraction and isolation

Desoxyerythrolaccin (**DE**) and **DDE-3-*O*-Glc*p*** were purified from 100 g of dried *D. coccus.* The insects were grinded and extracted with MeOH:H_2O (1:1) followed by liquid-liquid partitioning with EtOAc at pH 3 and concentrated *in vacuo*. The crude extract was first separated by ion-exchange using an NH_2 flash column. Carboxylic acid containing compounds was retained when washed with 50% aqueous MeCN containing 10 mM ammonium formate. Finally, the acidic compounds (**FK**, **KA**, and **CA**) were eluted with 50% aqueous MeCN adjusted to pH 11 with ammonium hydroxide. **DE** and **DDE-3-*O*-Glc*p*** were purified from the 50% MeCN 10 mM ammonium formate eluate using reversed-phase semi-preparative chromatography on a 250 × 10 mm Luna2 C_{18} column (Phenomenex, Torrance, CA, USA) using a Gilson HPLC system. Compounds were eluted with a gradient consisting of MilliQ H_2O:MeCN, both containing 50 ppm TFA. **FK** and **dcII** were recovered in the alkaline eluent of the NH2 column. The extract was then further purified on Isolute diol material (Biotage, Uppsala, Sweden) on an Isolera auto flash purification system (Biotage, Uppsala, Sweden) in a step-wise elution from dichloromethane to EtOAc to MeOH to afford **dcII** and **FK**.

2.4. Synthesis of flavokermesic acid anthrone

Synthesis of flavokermesic acid anthrone from flavokermesic acid was conducted according to a previously published method by Schätzle, with slight modifications (Schätzle, 2012). In brief, 10 mg flavokermesic acid was dissolved in 1mL glacial acetic acid and 0.2 mL hydriodic acid (57 wt. % in H_2O) in a sealed microwave reactions vial and heated to 50 °C under stirring for 2 h in the dark. This adaption to the method reported by Schätzle 2012 was done to ensure full conversion of the flavokermesic acid to the anthrone with only limited decarboxylation. The hydriodic acid was quenched with 10 mL saturated $Na_2S_2O_3$ and extracted three times with diethyl ether. The ether phase was dried over $MgSO_4$ and lyophilized under reduced pressure. The sample was at all times kept in the dark and chemical analyses were performed in amber vials and NMR tubes to minimize the risk of dimerization and oxidation. The formation of the anthrone was confirmed by NMR analysis (Supplementary data Table S1).

178

179 *2.5. Biosynthetic models*

180 The biosynthetic models for formation of CA and related compounds were formulated using the
181 retrosynthesis approach including commonly accepted enzymatic driven reactions, as described in the
182 BRENDA database (Schomburg et al., 2004), and the available structural data for coccid dyes (Morgan,
183 2010; Cameron et al., 1978, 1981; Brown, 1975, Peggie et al., 2007; Stathopoulou et al., 2013; Bhide et al.,
184 1969). The models were drawn using ChemDraw 15.9.9.106 (PerkinElmer Informatics, Inc., US).

185

186 **3. Results**

187 *3.1. Detection of compounds in Dactylopius coccus extracts*

188 HPLC-HRMS analysis of the raw extracts from fresh *D. coccus* showed that the main extractable pigment
189 components were **CA**, **FK**, **KA** and **dcII** as previously reported (Wouters, J., Verhecken, A., 1989; Peggie et
190 al., 2007; Méndez et al., 2004) (Figure 1A). These compounds were putatively identified based on relative
191 retention time, high-resolution mass, UV/VIS spectra, and MS/MS fragmentation patterns. Spiking with
192 authentic samples of **CA** and **KA** confirmed these compounds identity. **FK** was isolated and subjected to
193 structural elucidation by 2D NMR (Supplementary data Table S1), and the data were in agreement with and
194 confirmed the previously reported structure of **FK** (Wouters and Verhecken, 1987). **CA** and **dcII** were not
195 easily separated using reversed-phase HPLC, and isolation of the two compounds relied on normal-phase
196 flash chromatography on diol substituted silica. The putative **dcII** was further purified using semi-
197 preparative HPLC and characterized by HR-MS and 2D NMR. Structural elucidation of the compound by 2D
198 NMR experiments (Supplementary data Table S1) showed that **dcII** was flavokermesic acid 2-*C*-β-D-
199 glucopyranoside, and the NMR data were in agreement with those reported for **dcII** (Stathopoulou et al.,
200 2013). The reversed-phase HPLC-based analysis also revealed a previously undescribed major peak, eluting
201 at 13.40 minutes (Figure 1A and 1C). The mass of the corresponding compound equaled the theoretical
202 mass of flavokermesic acid anthrone (**FKA**), and this identity was confirmed by comparison with a **FKA**
203 standard, semi-synthesized from authentic **FK**. In addition, to the metabolites detected using reversed-
204 phase chromatography, strong anion exchange SPE (SAX SPE) were used to identify three non-acidic
205 metabolites (**DE**, **DDE-3-*O*-Glc*p* and DE-*O*-Glc*p***,) that all displayed UV/VIS spectra with similarities to those
206 reported for **FK**, **KA**, **dcII** and **CA** (Figure 1B, 1D, 1E and 1F). For unambiguously structure elucidation of **DE**
207 **and DDE-3-*O*-Glc*p***, the compounds were purified in amounts sufficient for structural elucidation by HRMS
208 and NMR spectroscopy. Purification of the **DE-*O*-Glc*p*** compound ufortionatly did not yield sufficient
209 quantities for full structural elucidation by NMR.

210

*3.2. Structural elucidation of **DE**, and the novel compounds **DE-O-glucosyl**, **DDE-O-glucosyl** and **FKA***

DE displayed UV/VIS spectrum (Figure 1E) similar to that of **FK**, indicating a similar core skeleton. The compound was not retained on a SAX column, suggesting that it lacked the carboxylic acid group found at C-7 in **FK**. This conclusion was supported by HRMS (m/z 271.0600 $[M+H]^+$, calcd 271.0600, ΔM 0.0 ppm), suggesting a molecular formula of $C_{15}H_{10}O_5$, i.e., **DE** lacking CO_2 as compared to **FK**. Structural elucidation was carried out by ^{1}H NMR and 2D NMR spectroscopy (Table S1). The ^{1}H NMR spectrum showed a signal for the OH-group positioned *peri* to the carbonyl group (δ 13.30, 1-OH), two sets of meta-coupled protons H-5 and H-7 (δ 6.54 and 7.04, respectively, $^3J_{H5\text{-}H7}$ = 1.9 Hz) and H-2 and H-4 (δ 7.43 and 7.01, respectively, $^3J_{H2\text{-}H4}$ = 2.5 Hz), and a *peri*-positioned methyl group (δ 2.81, s, 11-CH_3). The meta-coupling between H-5 and H-7 clearly proves the lack of the carboxylic acid in position 7, and thus HRMS and NMR data supported the compound to be **DE**, also known as 3-hydroxy-aloesaponarin II (Mehandale et al., 1968), and previously observed in air-dried *D. coccus* (Sugimoto et al., 1998).

Structural analysis of the first novel compound **DDE-3-*O*-Glc** revealed that the compound was a hitherto undescribed *O*-glucoside of 5,6-didehydroxyerythrolaccin (**DDE**) (Figure 2). The high-resolution mass spectrum of **DDE-3-*O*-Glc*p*** suggested a molecular formula of $C_{21}H_{20}O_9$ ($[M+H]^+$ *m/z* 417.1180, calcd. $C_{21}H_{21}O_9^+$ 417.1180, ΔM 0.0 ppm; $[M-H]^-$ *m/z* 415.1034, calcd. $C_{21}H_{19}O_9^-$ 415.1029, ΔM 1.2 ppm. In addition, the compound exhibited a loss of *m/z* 162.0528, which is likely due to the loss of a labile *O*-linked hexose unit. The structural elucidation was carried out based on ^{1}H NMR and 2D NMR spectroscopy (Supplementary data Table S1). The ^{1}H spectrum showed resemblance to that of **DE**, but instead of the two doublets observed for the meta-coupled H-5 and H-7 in **DE**, signals for H-5 (δ 8.12, dd, 7.5, 1.2 Hz), H-6 (δ 7.78, t, 7.5 Hz), and H-7 (δ 7.75 dd, 7.5, 1.3 Hz) showed the absence of a hydroxyl group at C-6 in **DDE-3-*O*-Glc*p***. Furthermore, a doublet at δ 5.26 ($^3J_{H1'\text{-}H2'}$ = 7.6 Hz) for a β-configuration of the anomeric proton H-1' (as well as the ^{13}C value of 101.1 ppm for C-1) and the remaining ^{1}H and ^{13}C signals for H-2' to H-6' and C-2' to C-6' (Supplementary data Table S1), are in agreement with a β-D-glucose unit (Bock and Pedersen, 1983). The *O*-linkage of the sugar was evident from a more deshielded anomeric proton (101.1 ppm) compared to that of *C*-glucosyl linkages reported for related compounds (Stathopoulou et al., 2013). In addition, a HMBC correlation from H-1' to C-3 (164.6 ppm) further confirmed the *O*-glucosyl linkage to C-3. Thus, to establish the D- or L-configuration of the glucose moiety, an aliquot was hydrolyzed and analyzed by GC-MS and chiral GC-MS. This confirmed that the hexose moiety was D-glucose (Figure S2 and S3). Thus, the compound was identified as 5,6-didehydroxyerythrolaccin 3-*O*-β-D-glucopyranoside.

Several attempts to purify the second novel compound **DE-*O*-Glc*p*** did not yield sufficient quantities to allow acquisition of NMR data. However, the high-resolution mass spectrum of **DE-*O*-Glc*p*** suggested a molecular formula of $C_{21}H_{20}O_{10}$ ($[M+H]^+$ *m/z* 433.1129, calcd. $C_{21}H_{21}O_{10}^+$ 433.1129, ΔM 0.0 ppm; $C_{21}H_{20}O_{10}$

([M+H]$^+$ *m/z* 431.0981, calcd. $C_{21}H_{19}O_{10}^-$ 431.0984, ΔM 0.7 ppm); and the loss of a hexose moiety (*m/z* 162.0530) generated a fragment with the same mass as **DE**. These results indicate that this compound is an *O*-glucosylated form of **DE** (Figure 2), but the exact position of glucosylation could not be established.

The third novel compound was only detected in extracts from fresh and frozen *D. coccus* and identified as flavokermesic acid anthrone (**FKA**) (Figure 2). The high-resolution mass spectrum of **FKA** detected in the insect suggested a molecular formula of $C_{16}H_{12}O_6$ ([M-H]$^-$ *m/z* 299.0559, calcd. $C_{16}H_{11}O_6^-$ 299.0561, ΔM 1.0 ppm). Positive identification of this compound was achieved by comparison of retention time, high-resolution MS, MS/MS fragmentation pattern, and UV/VIS spectrum for **FKA that had been prepared by chemical semi-synthesis**. See Table S1 for NMR data. Chemical semi-synthesis of **FKA** demonstrated that the pure compound is prone to dimerization as well as oxidation to **FK** in the presence of oxidants under *in vitro* conditions. The observed spontaneous oxidation to **FK** likely explains why only fresh and frozen *D. coccus* was found to contain **FKA** while the compound was not detected in dried insects which have been exposed to light, oxygen and other oxidative agents for longer periods of time.

3.3. LC-DAD/MS-based screening of coccid dye production in selected members of the superfamily Coccoidea

To analyze the taxonomic distribution of the ability to produce coccid dyes, we collected representatives of four different *Coccoidea* families found in Denmark. The collected species included nymphal states of *Coccus hesperidium*, *Pseudococcus longispinus*, *Palmicultor browni* and *Pseudaulacaspis pentagona*. Metabolites from approximately 1 g of each of the individual species were extracted and analyzed by LC-MS/DAD. Extracted ion chromatograms (Figure 3) for masses equivalent to the known coccid dye intermediates showed that *C. hesperidium* contained **KA**, **FK**, **dcII** and **CA**, while *P. longispinus* and *P. browni* both contained **FK** and **dcII**. Positive identification was based on accurate mass, retention time, UV/VIS spectra and authentic standards. Analysis of the *P. pentagona* material did not reveal the presence of any coccid dye intermediates in this species.

4. Discussion

Assyrian cuneiform texts document that the utilization of scale insects for the production of pigments and dyes dates back to at least 3000 BC (Dapson 2007). Different civilizations separated by time and geography have relied on a limited number of scale insect species and the pigments they produce (Morgan, 2010; Cameron et al., 1978, 1981; Brown, 1975, Petitpierre, 1981: Peggie et al., 2007; Stathopoulou et al., 2013; Bhide et al., 1969). The long history and traditions associated with coccid dye utilization have resulted in a situation where scientific literature has focused on a very limited number of producing species (*Dactylopius*

coccus, *Kerria lacca,* and *Kermes vermilio*) indicating a discontinuous taxonomic distribution of the trait and unrelated biosynthetic origins for the various pigments. However, the number of shared metabolites and the common **FK** core structure of the pigments suggest that all coccid dyes are formed via a similar biosynthetic scheme with a common evolutionary origin. If so, then many more scale insect species, i.e. those forming a monophyletic clade with known producers, would also be expected to be able to produce coccid dyes or related metabolites. To this end, we have demonstrated that members of the *Planococcus* (mealybugs) and *Pseudaulacaspis* genera also produce coccid dyes. Hence, the ability to form these pigments is indeed more widely taxonomically spread than previously believed. This would support the most parsimonious explanation, being a common evolutionary origin within Coccoidea. Of evolutionary importance here is that *Porphyrophora* and *Dactylopius* belong to two vastly different clades of Coccoidea. *Porphyrophora* belongs to the more primitive Margarodidae and *Dactylopius* to the Dactylopiidae within the more highly derived Neococcoidea clade. The Margarodidae and Dactylopiidae share a most recent common ancestor 250 million years ago (Vea et al., 2016).

4.1. The biosynthetic origin of coccid dyes

The biosynthetic origin of coccid dyes such as **KA**, **FK**, **CA**, and **LA** has long been debated (Brown, 1975; Joshi and Lambdin, 1996; Ramirez-Puebla et al., 2010). Several studies have rejected the hypothesis that host plants supply the insects with any of the known coccid dye intermediates. Similarly, we were unable to detect any of these intermediates in *Opuntia* cacti pads in the case of *D. coccus* (data not shown). To say nothing about the wide variety of different hosts utilized by many *Coccoidea*, e.g. over 400 plant species are described as host for *Kerria lacca* (Sharma et al., 1997).

Accordingly, it seems clear that Coccoidea must be able to synthesize the coccid dyes *de novo* from simple metabolites, e.g. glucose, present in the phloem sap of a wide range of plant species. Several studies have suggested that endosymbiotic bacteria may be responsible for the formation of the coccid dyes, and members of the *Wolbachia* and *Azoarcus* bacteria genera have been identified in the scale insects (Brown, 1975; Ramirez-Puebla et al., 2010; Pankewitz et al., 2007). However, none of these studies have proven a direct link between the presence, or activity, of these endosymbionts with the formation of coccid dyes. An alternative hypothesis is that the biosynthetic apparatus is encoded in the insects' nuclear genome. Several examples exist where complex secondary metabolite's biosynthetic pathways are encoded by genes in the genome of the producing insect, such as the *Drosophila* eye pigments drosopterin and ommochromes (Chovnick et al., 1990; Nijhout, 1997). Though no one has yet succeeded in identifying or describing the enzymological- or genetic basis for polyketide biosynthesis in insects, strong evidence does exist in support of nuclear encodement of the enzymes required for producing coccid dyes. However, this has largely been

overlooked in past studies. Such evidence includes the report of a stable yellow color mutant of the normally red *Kerria lacca* (Indian Lac-insect) (Negi, 1954), and a white color mutant incapable of producing **LA**s (Chauhan and Teotia, 1973). Dissection of the genetic basis for these two mutations, by classical genetic crossing experiments, showed that the two traits are non-allelic and that they follow simple recessive inheritance (Chaucun, 1977; Chaucun and Mishra, 1977). If endosymbiotic bacteria were involved in catalyzing steps in the formation of the core structure of the coccid dyes, the mutant trait would be expected to be maternally inherited as a result of transfer via eggs (Ferrari and Vavre, 2011). Any conclusions on this matter must await demonstration of the ability of the color mutants to host endosymbiotic bacteria. Independently, support of an insect rather than bacterial origin of **CA** has been provided by a series of microscopy studies of the hemolymph from various *Dactylopius* spp. These studies described the existence of special granulocytes with a high concentration of secretory (M-) granules containing red pigments and small corpuscles of **CA** floating freely in the hemolymph (Joshi and Lambdin, 1996; Caselin-Castro et al., 2008, 2010).

4.2. Model for the biosynthesis of carminic acid

Based on its structure, the biosynthesis of **CA** may be hypothesized to proceed by two different routes (Figure 4). One envisioned biosynthetic scheme involves the polyketide pathway, also known as the acetate/malonate pathway (Figure 4A). A second possible biosynthetic route involves the shikimate based chorismate/*O*-succinyl benzoic acid pathways (Figure 4B). Both biosynthetic schemes ultimately result in the formation of anthraquinones; however they would be predicted to include different intermediates, e.g. a unique anthrone in the case of a polyketide-based pathway, which can be used to distinguish between the two. Detection of the **FK** anthrone (**FKA**) in fresh and frozen *D. coccus* material (Figure 1) provides support for a polyketide rather than shikimate origin of compounds with an **FK** core. That the anthrone is detected in the current study can likely be attributed to the milder extraction conditions, and the use of fresh material rather than dried insects or dyed textiles that have been used in previous reports on the subject. The anthrone is abundant in the fresh material as evident by the HPLC-HRMS-DAD analysis (Figure 1), but is not detected in dried insects, which is in good agreement with our previous observation that the purified anthrone spontaneous oxidizes. Microbial based reduction of the **FK** anthraquinone to yield the **FKA** anthrone is a possible alternative explanation for detection of the anthrone. de Witte and co-workers have previously shown that bacteria isolated from mammalian fecal material are capable of catalyzing the reduction of the anthraquinone rhein to the corresponding anthrone. The currently available data do not allow us to rule out this explanation, and further experiments e.g. feeding experiments are hence required (de Witte et al. 1992).

Based on the detection of FKA and the novel **CA** related compounds identified in *D.* coccus, and under the assumption that its formation is not the result of microbial reduction, we here propose a biosynthetic pathway for the formation of **CA** in *D. coccus* (Figure 5). This biosynthetic pathway is an elaboration of the models previous proposed by Brown (Brown, 1975) and Morgan (Morgan, 2010), and differs by including additional intermediates and predictions for the required enzymatic activities and co-factors.

The enzymatic machinery, responsible for the formation of polyketides in animals, remains unknown, and several competing hypotheses exists. One possible explanation could be that the involved PKS has been introduced into the genome of scale insects by horizontal gene transfer (HGT) from fungi or bacteria. Several examples of HGT from fungi to insects have previously been documented such as the carotenoid forming pathway in *Acyrthosiphon pisum* (pea aphid) (Moran and Jarvik, 2010). Synthesis of the **FK** core requires the formation of a C7-C12 intermolecular bond in the octaketide backbone (Figure 5). Since fungal type I iterative PKSs have only been described to form either C2-C7 or C6-C11 bonds, fungi are an unlikely donor (Li et al., 2010). Engineered bacterial type II PKS systems have previously been shown to be able to produce **FK**, known as **TMAC** in the bacterial literature (Tang et al., 2004). Specifically, **TMAC** is formed by the combined actions of the minimal actinorhodin PKS (act-KSα, act-KSβ, act-ACP) from *Streptomyces coelicolor* and the two cyclases (ZhuI and ZhuJ) from *Streptomyces* sp. No. 1128 (Tang et al., 2004). To settle whether HGT has formed the basis for **CA** production requires that the responsible genes are identified and analyzed in the context of a high-quality scale insect genome sequence, which is not presently available.

A competing hypothesis for the origin of insect PKSs, and the one we favor, is that the putative PKSs may have evolved from the insect's endogenous type I fatty acid synthase (FAS). Animal type I FAS and fungal type I iterative PKS are thought to have evolved from a common bacterial type I PKS ancestor (Hopwood and Sherman, 1990; Kroken et al., 2003). The main product of FASs in animals is palmitic acid, a fully reduced C-16 chain. This chain length is equivalent to an octaketide, which is the intermediate required for **FKA** formation. Converting a FAS to a non-reducing PKS capable of producing a non-reduced linear octaketide would require inactivation of the FAS's β-ketoreductase (KR) domain combined with a relaxation of the substrate specificity of its β-ketosynthase domain (KS) to allow for non-reduced products to form. Non-reduced linear polyketides are highly reactive, due to the presence of carbonyl groups on every second carbon atom, and they spontaneously fold into heterocyclic and aromatic structures via the formation of intramolecular C-C bonds. Several studies have shown that non-reduced octaketides spontaneously form the aromatic compounds **SEK4** and **SEK4b**, which contain intermolecular bond configurations that differ significantly from that of **FKA** (Figure 5) (Fu et al., 1994; Mizuuchi et al., 2009). The fact that we find only **FKA**, and not **SEK4** or **SEK4b**, in *D. coccus* extracts (data not shown) suggests that folding of the polyketide chain does not proceed as a spontaneous reaction. Folding control of non-reduced polyketide backbones in

fungal type I iterative non-reducing PKS systems and bacterial type III PKS systems are achieved by a 'Product Template' (PT) domain in the PKS (evolved from DH domain) or by *trans*-acting cyclases and aromatases, respectively (Shen et al., 1995; Bringmann et al., 2006). In the case of coccid pigments, controlled folding of the linear octaketide to form **FKA** could depend on similar mechanisms that would require additional mutations in the FAS or trans-acting enzymes as presented in Figure 5. It has not escaped our notice that the mutated FAS hypothesis potentially also can explain the formation of other polyketides found in insects, such as 5-hydroxy-7-methyl-6-acetylpurpurin from *Ericoccus* spp. (Coccoidea: Ericoccidae) (Banks and Cameron, 1970), chrysophanol in *Galeruca tanaceti* (Coleoptera: Chrysomelidae) (leaf beetle) (Bringmann et al., 2006), and the predicted monomeric precursors of protoaphins in aphids (Brown, 1975). The listed compounds are all likely also formed from non-reduced octaketide precursors but display alternative backbone folds and would hence dependent on other cyclases than those involved in coccid dyes biosynthesis.

In the case of **CA** formation, the enzymatic steps following formation of **FKA** are predicted to include two oxidations and a *C*-glucosylation. Based on the metabolites detected in *D. coccus,* monooxygenation of the central aromatic ring (position C-10), from **FKA** to **FK**, likely occur before oxidation of the outer ring (position C-4) and before *C*-glucosylation (position C-2), as neither of the detected metabolites contain a C-4 oxidation without a C-10 oxidation and as all known glucosylated intermediates (e.g. dcII and CA) have the C-10 oxidation. Several studies of bacterial and fungal systems have shown that efficient *in vivo* anthrone oxidation is dependent on specific anthrone oxidases (Chung et al., 2002; Erhlich et al. 2010), making it likely that **FKA** to **FK** conversion is an enzyme dependent reaction *in vivo*, rather than a spontaneous reaction. The order of the two subsequent reactions (monooxygenation of C-4 and C-glucosylation at C-2) that ultimately yield **CA** is unclear as both **KA** and **dcII** accumulates. Hence, it is impossible to determine whether both pathways are active *in vivo* or whether one represents a shunt. Monooxygenation of the C-4 position that converts **FK** to **KA** is likely catalyzed by either a cytochrome P-450 or flavin-dependent monooxygenases. The accumulation of several intermediates suggests that the natural **CA** biosynthetic pathway is imbalanced; a situation that likely is caused by insufficient flux through downstream enzymatic steps resulting in the buildup of intermediates.

4.3. Decarboxylation of the FKA core

We also demonstrated the presence of **DE**, **DE-*O*-Glc*p*,** and **DDE-3-*O*-Glc*p*** in freshly collected insects. **DE** has previously been observed in air-dried *D. coccus* (Sugimoto et al., 1998). Dehydroxy- and *O*-glucosylated forms of **DE** had not previously been reported in *D. coccus*. The reason why we detect the *O*-glucosylated forms may be attributed to the mild extraction conditions and moderate pH compared to previously

reported extraction protocols, which may result in hydrolysis of *O*-glucosides. *Kerria laccas* (lac insect) and *Austrotachardia acacia* (Maskell) are also known to accumulate **DEL**, erythrolaccin (**EL**) and iso-erythrolaccin (**IEL**) in their resin deposits (Chauhan, 1977; Caselin-Castro et al., 2010). The structural similarity and co-occurrence in multiple species suggest a common biosynthetic origin for the **FK** (C16) and **EL** (C15) compound families. This is further supported by the observation made by Chauhan and Mishra (Chauhan and Mishra, 1977) who noted that a single mutation in white *K. lacca* strains affected both the body color (primarily caused by **LA**) and the resin color (primarily caused by **EL**) (Bhide et al., 1969). Based on this, we propose that the **FK** (C16) and **EL** (C15) compound families are products of the same biosynthetic mechanism in scale insects and that the difference in carbon number is due to decarboxylation of **FKA**, as presented in Figure 5. A highly similar anthraquinone decarboxylation step has previously been documented in the chrysophanol biosynthetic pathway in *G. tanaceti*, though the responsible mechanism and timing of the decarboxylation step is unknown (Bringman et al., 2006).

4.4. Evolution of the FK biosynthetic pathways / Variations to the FK biosynthetic pathway

CA has been shown to act as a chemo deterrent that protects the immobile scale insects from predatory ants (Eisner et al., 1980). However, García Gil de Muñoz and co-workers recently extended **CA**'s biological function by showing that it may contribute to the innate immune system of the scale insect to protect against invading microorganisms (García-Gil De Muñoz et al, 2002, 2005; De La Cruz Hernandez-Hernandez, 2003; García-Gil De Muñoz, 2007). Specifically, this system depends on encapsulation of the invaders by melanization via the rapid formation of eumelanin by polymerization of tyrosine and L-DOPA (Satyavathi et al., 2014; Charles and Killian, 2015). Phenoloxidase (PO) is responsible for catalyzing multiple steps of the melanization cascade: tyrosine to L-DOPA, L-DOPA to dopaquinone, dopamine to dopaminequinone, and dopamine to *N*-arachidonoyl dopamine. Reactions that generate reactive radicals, including reactive oxygen species (ROS) and DOPA semi-quinones, which in addition to encapsulation may harm intruders (González-Santoyo and Córdoba-Aguilar, 2012). As shown by Garcia *et al.*, PO can also act directly on **CA** resulting in the formation of insoluble polymers of **CA** (García-Gil De Muñoz et al, 2005; García-Gil De Muñoz, 2007). This process is attributed to the quinone nature of **CA** allowing it to participate in redox cycling with ROSs. Relocation of the radical within the conjugated system of **CA** or **FK** would allow for activation of multiple positions in the **FK** core (C4, C8, C10, and C9-OH) as described for other naphthoquinones (Frandsen et al., 2006). This is a very interesting observation as the formation of **FK** radicals potentially can explain how the tyrosine-derived groups found in **LAs** are added to the **FK** core by radical activation of the C10 positions and oxidative coupling with tyrosine or one of its derivates (tyrosol, N-acetyltyramine, tyramine or 2-(4-hydroxyphenyl)ethyl acetate) (Figure 6). Figure 6 summarizes the chemical diversity and the required

enzymatic steps in the form of a meta-biosynthetic pathway accounting for all known coccoid pigments. The model includes five different monooxygenases, acting on C4, C6, C8, C16, two dehydrogenases, a decarboxylase and a *C*-glucosyltransferase.

The extensive chemical diversity and existence of multiple alternative decoration patterns support the hypothesis that the **FK** forming biosynthetic pathway has a long evolutionary history within the *Coccoidea* superfamily.

In summary, we propose that formation of **CA** depends on the activity of a modified fatty acid synthase or polyketide synthase, possibly one or more cyclases/aromatases, one anthrone oxidases, a 'cytochrome P450 monooxygenases'/'flavin-dependent monooxygenases', and a *C*-glucosyl transferase. Validation of the proposed hypothetical biosynthetic schemes and the involved enzyme types naturally depends on future biochemical evidence and mapping of their genetic basis in *D. coccus* or an endosymbiont organism.

Abbreviations

Flavokermesic acid anthrone (**FKA**); flavokermesic acid (**FK**); kermesic acid (**KA**); carminic acid (**CA**); C-glucosylated flavokermesic acid (**dcII**); laccaic acid (**LA**); desoxyerythrolaccin (**DE**), 5,6-didehydroxyerythrolaccin (**DDE**); didehydroxyerythrolaccin 3-*O*-β-D-glucopyranoside (**DDE-3-*O*-Glc*p***); erythrolaccin (**EL**); iso-erythrolaccin (**IEL**)

Acknowledgements

The Danish Innovation Foundation is acknowledged for funding the work via grant no. XXXXXXXX. Center for Synthetic Biology funded by the University of Copenhagen's Excellence Program for Interdisciplinary Research is acknowledged for technical and financial support. The US NSF for funding AVD's involvement in the project via grant no. DBI-1306489. We would like to thank Cultivo de la cochinilla en Mala y Guatiza for their assistance with collection the live *Dactylopius coccus* specimens on Lanzarote. We are also thankful for Jimmy Oluf Olsen, from the Royal Botanical Garden, the Natural History Museum of Denmark, University of Copenhagen, for his assistance with collecting.

Conflicts of interest

Authors SAR, KTK, DS, CHG, UT, UHM, TOL, and RJNF declare no financial nor any competing financial interests. The authors, PKJ, MB and BM are or were employed by the private company Chr. Hansen A/S that produces and sells *D. coccus* derived carmine as a food pigment in a business-to-business setup.

475 **Appendix A. Supplementary data**

476 Supplementary data related to this article can be found at http://dx.doi.org/ToBeInserted

477

478 **References**

479 Banks, H., Cameron, D., 1970. Polyhydroxyanthraquinones from the insect *Eriococcus coriaceus*. J. Chem.
480 Soc. D 22, 1577-1578

481 Bhide, N., Pandhare, E., Rama Rao, A., Shaikh, I., Srinivasan, R., 1969. Lac pigments: Part IV. Constitution of
482 laccaic acid B. Indian J. Chem. 7, 987–995

483 Bock, K. Pedersen, C., 1983. Advances in Carbohydrate Chemistry and Biochemistry Volume 41 (Tipson, S.
484 R., and Horton, D. eds), Advances in Carbohydrate Chemistry and Biochemistry

485 Bringmann, G., Noll, T.F., Gulder, T.A.M., Grüne, M., Dreyer, M., Wilde, C., Pankewitz, F., Hilker, M., Payne,
486 G.D., Jones, A.L., Goodfellow, M., Fiedler, H.-P., 2006. Different polyketide folding modes converge to an
487 identical molecular architecture. Nat. Chem. Biol. 2, 429–433

488 Brown, K., 1975. The chemistry of aphids and scale insects. Chem. Soc. Rev. 4, 11–36

489 Cameron, D.W., Feutrill, G.I., Karga, R.L., Patti, A.F., 1981. Chemistry of the Coccoidea. IX. Pigments from
490 the lac insects *Austrotachardia acaciae* (Hemiptera) and associated synthetic work. Aust. J. Chem. 34,
491 2481–2485

492 Cameron, D., Feutrill, G., Patii, A., Raston, C., Raverty, W., White, A., 1978. Chemistry of the Coccoidea. VI.
493 Pigments of the lac insect *Austrotachardia acaciae* (Hemiptera): chemistry and crystal structure. Aust. J.
494 Chem. 31, 2651–2658

495 Caselin-Castro, S., Llanderal-Cazares, C., Ramirez-Cruz, A., Hernandez, M.S., Mendez-Montiel, J.T., 2008.
496 Morphological characterization of hemocytes from the female *Dactylopius coccus* Costa
497 (Hemiptera:Coccoidea:Dactylophdae). Agrociencia 42, 349–355

498 Caselín-Castro, S., Llanderal-Cázares, C., de Jesús Méndez-Gallegos, S., Ramírez-Cruz, A., de la Cruz
499 Hernández-Hernández, F., 2010. Hemocytes of the cochineal insect: Ultrastructure. Arch. Insect
500 Biochem. Physiol. 73, 176–192

501 Charles, H., Killian, K., 2015. Response of the insect immune system to three different immune challenges. J
502 Insect Physiol. 81, 97–108

503 Chauhan, N., Teotia, T., 1973. A new variant in the lac insect, *Kerria lacca*. Entomol Newslett. 3, 33–34

504 Chauhan, N., 1977. Gene expression and transmission in *Kerria lacca* (Kerr). Heredity (Edinb). 38, 155–159

505 Chauhan, N., Mishra, Y., 1977. White: A new colour locus in *Kerria lacca* (Kerr). Curr. Sci. 46, 272–273

506 Chovnick, A., Clark, S. H., Dutton, F. L., Reaume, A. G., 1990. The rosy locus and xanthine dehydrogenase in
507 *Drosophila melanogaster*. Prog. Clin. Biol. Res. 344, 1–18

508 Chung, J. Y., I. Fujii, S. Harada, U. Sankawa, Ebizuka Y., 2002. Expression, purification, and characterization
509 of AknX anthrone oxygenase, which is involved in aklavinone biosynthesis in *Streptomyces galilaeus*. J.
510 Bacteriol. 184:6115-6122

511 Dapson, R.W., 2007. The history, chemistry and modes of action of carmine and related dyes. Biotech.
512 Histochem. 82, 173–187

513 De La Cruz Hernandez-Hernandez, F., García-Gil De Muñoz, F., Rojas-Martínez, A., Hernández-Martínez, S.,

514 Lanz-Mendoza, H., 2003. Carminic acid dye from the homopteran *Dactylopius coccus* hemolymph is
515 consumed during treatment with different microbial elicitors. Arch. Insect Biochem. Physiol. 54, 37–45

516 de Witte, P., van Hoestenberghe, A., Eyssen, H., 1992. *In vitro* reduction of rhein anthraquinone to rhein
517 anthrone by rat cecal microflora and some intestinal bacterial strains. Biopharm Drug Dispos. 13(4):243-
518 53.

519 Donkin, R.A., 1977. The insect dyes of Western and West-Central Asia. Anthropos 72, 847–880

520 Dooley, J., Dones, R., 2015. Armored scales of the World: Compendium and key to the genera of the
521 Diaspididae (Sternorryncha, Hemiptera).
522 http://*keys.lucidcentral.org/keys/v3/Dones_Lourdes/homepage.htm*

523 Downham, A., Collins, P., 2000. Colouring our foods in the last and next millennium. Int. J. Food Sci.
524 Technol. 35, 5–22

525 Ehrlich K.C., Li P, Scharfenstein L., Chang P.K., 2010. HypC, the anthrone oxidase involved in aflatoxin
526 biosynthesis. Appl Environ Microbiol. 76(10):3374-7

527 Eisner, T., Nowicki, S., Goetz, M., Meinwald, J., 1980. Red cochineal dye (carminic acid): its role in nature.
528 Science 208, 1039–1042

529 Eisner, T., Ziegler, R., McCormick, J.L., Eisner, M., Hoebeke, E.R., Meinwald, J., 1994. Defensive use of an
530 acquired substance (carminic acid) by predaceous insect larvae. Experientia 50, 610–615

531 Ferrari, J., Vavre, F., 2011. Bacterial symbionts in insects or the story of communities affecting communities.
532 Philos. Trans. R. Soc. Lond. B. Biol. Sci. 366, 1389–1400

533 Frandsen, R.J.N., Nielsen, N.J., Maolanon, N., Sørensen, J.C., Olsson, S., Nielsen, J., Giese, H., 2006. The
534 biosynthetic pathway for aurofusarin in *Fusarium graminearum* reveals a close link between the
535 naphthoquinones and naphthopyrones. Mol. Microbiol. 61, 1069–1080

536 Fu, H., Hopwood, D.A., Khosla, C., 1994. Engineered biosynthesis of novel polyketides: evidence for
537 temporal, but not regiospecific, control of cyclization of an aromatic polyketide precursor. Chem. Biol. 1,
538 205–210

539 García-Gil De Muñoz, F., 2007. Free radical generation during the activation of hemolymph prepared from
540 the homopteran *Dactylopius coccus*. Arch Insect Biochem. Physiol. 28, 20–28

541 García-Gil De Muñoz, F., Calzada, A., Rangel, C., Castro, G., Lanz, H., del Río, I., Hernandez, F. (2005)
542 Generación de anión superóxido O2- en respuesta a zymosan (β, 1-3 glucanos), activador de la respuesta
543 inmune en la cochinilla fina del nopal (Dactylopius coccus: Homóptera). Fac. Cienc. y Tecnol. Investig.
544 Univ. Multidiscip. 4, 26–31

545 García-Gil De Muñoz, F., Lanz, H., Rojas, A., Hermandez, F., 2002. Efecto de inhibidores de la sintesis de
546 prostaglandinas en la coagulacion del homptero *Dactylopius coccus* (cochinilla del nopa). Investifacion
547 Univ. Multidiscip. Rev. Investig. la Univ. Simon Boliv. 1, 14–19

548 González-Santoyo, I., Córdoba-Aguilar, A., 2012. Phenoloxidase: A key component of the insect immune
549 system. Entomol. Exp. Appl. 142, 1–16

550 Han, Y.S., Van Der Heijden, R., Verpoorte, R., 2001. Biosynthesis of anthraquinones in cell cultures of the
551 Rubiaceae. Plant Cell. Tissue Organ Cult. 67, 201–220

552 Hopwood, D.A., Sherman, D.H., 1990. Molecular genetics of polyketides and its comparison to fatty acid
553 biosynthesis. Annu. Rev. Genet. 24, 37–66

554 Horikawa, M., Hoshiyama, T. S., Matsuzawa, M., Shugyo, T., Tanaka, M., Suzuki, S., Sato, M., Ito, T.,

555 Asakawa, Y., Kaku, H., Nishii, T., Inai, M., Takahashi, S., Tsunoda, T., 2011. Viridaphin A 1 glucoside, a
556 green pigment possessing cytotoxic and antibacterial activity from the aphid *Megoura crassicauda*. J.
557 Nat. Prod. 74, 1812–1816

558 Joshi, P.A., Lambdin, P.L., 1996. The ultrastructure of hemocytes in *Dactylopius confusus* (Cockerell), and
559 the role of granulocytes in the synthesis of cochineal dye. Protoplasma 192, 199–216

560 Kellner, R.L.L., 2002. Molecular identification of an endosymbiotic bacterium associated with pederin
561 biosynthesis in *Paederus sabaeus* (Coleoptera: Staphylinidae). Insect Biochem. Mol. Biol. 32, 389–395

562 Kroken, S., Glass, N.L., Taylor, J.W., Yoder, O.C., Turgeon, B.G., 2003. Phylogenomic analysis of type I
563 polyketide synthase genes in pathogenic and saprobic ascomycetes. Proc. Natl. Acad. Sci. U S A 100,
564 15670–15675

565 Łagowska, B., Golan, K., 2009. Scale insects /Hemiptera, Coccoidea/ as a source of natural dye and other
566 useful substances. Aphids and other Hemipterous insects 15, 151–167

567 Leistner, E., 1973. Biosynthesis of morindon and alizarine in intact plants and cell suspension cultures of
568 *Morinda citrifolia*. Phytochemistry 12, 1669–1674

569 Li, Y., Image, I.I., Xu, W., Image, I., Tang, Y., Image, I., 2010. Classification, prediction, and verification of the
570 regioselectivity of fungal polyketide synthase product template domains. J. Biol. Chem. 285, 22764–
571 22773

572 Mehandale, A.R., Rama Rao, A.V., Shaikh, I.N., Vankataraman K., 1968. Desoxyerythrolaccin and laccaic acid
573 D. Tetrahedron Lett. 18, 2231–2234

574 Méndez, J., González, M., Lobo, M.G., Carnero, A., 2004. Color quality of pigments in cochineals
575 (*Dactylopius coccus* Costa). Geographical origin characterization using multivariate statistical analysis. J.
576 Agric. Food Chem. 52, 1331–1337

577 Miller, D.R., Rung, A, Parikh, G., 2014. Scale Insects, edition 2, a tool for the identification of potential pest
578 scales at U.S.A. ports-of-entry (Hemiptera, Sternorrhyncha, Coccoidea). Zookeys 431, 61–78

579 Mizuuchi, Y., Shi, S.P., Wanibuchi, K., Kojima, A., Morita, H., Noguchi, H., Abe, I., 2009. Novel type III
580 polyketide synthases from *Aloe arborescens*. FEBS J. 276, 2391–2401

581 Moran, N.A., Jarvik, T., 2010. Lateral transfer of genes from fungi underlies carotenoid production in aphids.
582 Science 328, 624–627

583 Morgan, E.D., 2010. *Biosynthesis in Insects, Advanced Edition*, Royal Society of Chemistry, pp. 275-276,
584 Cambridge,

585 Negi, P.S., 1954. Dimorphism in lac insects. Indian For. 80, 423–426

586 Nijhout, H. F. (1997) Ommochrome pigmentation of the *linea* and *rosa* seasonal forms of *Precis coenia*
587 (Lepidoptera: Nymphalidae). Arch. Insect Biochem. Physiol. 36, 215–222

588 Pankewitz, F., Hilker, M., 2008. Polyketides in insects: ecological role of these widespread chemicals and
589 evolutionary aspects of their biogenesis. Biol. Rev. 83, 209–226

590 Pankewitz, F., Zöllmer, A., Hilker, M., Gräser, Y., 2007. Presence of Wolbachia in insect eggs containing
591 antimicrobially active anthraquinones. Microb. Ecol. 54, 713–721

592 Peggie, D.A., Hulme, A.N., McNab, H., Quye, A., 2007. Towards the identification of characteristic minor
593 components from textiles dyed with weld (*Reseda luteola* L.) and those dyed with Mexican cochineal
594 (*Dactylopius coccus* Costa). Microchim. Acta. 162, 371–380

595 Petitpierre, E., 1981. Presence of anthraquiones in the haemolymph of *Timarcha*. Chrysomela 30, 4

596 Piel, J., 2002. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont
597 of *Paederus* beetles. Proc. Natl Acad. Sci. U S A 99, 14002–14007

598 Ramirez-Puebla, S.T., Rosenblueth, M., Chavez-Moreno, C.K., Catanho Pereira de Lyra, M.C., Tecante, A.,
599 Martinez-Romero, E., 2010. Molecular phylogeny of the genus *Dactylopius* (Hemiptera: Dactylopiidae)
600 and identification of the symbiotic bacteria. Environ. Entomol. 39, 1178–1183

601 Satyavathi, V.V., Minz, A., Nagaraju, J., 2014. Nodulation: An unexplored cellular defense mechanism in
602 insects. Cell. Signal. 26, 1753–1763

603 Schätzle, M.A., Husain, S.M., Ferlaino, S., Müller, M., 2012. Tautomers of anthrahydroquinones: enzymatic
604 reduction and implications for chrysophanol, monodictyphenone, and related xanthone biosyntheses. J.
605 Am. Chem. Soc. 134, 14742–14745

606 Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., Schomburg, D., 2004. BRENDA, the
607 enzyme database: updates and major new developments. Nucleic Acids Res. 32, D431–D433

608 Shen, B., Summers, R., Wendt-Pienkowski, E., Hutchinson, C., 1995. The *Streptomyces glaucescens tcmKL*
609 polyketide synthase and *tcmN* polyketide cyclase genes govern the size and shape of aromatic
610 polyketides. J. Am. Chem. Soc. 117, 6811–6821

611 Stathopoulou, K., Valianou, L., Skaltsounis, A.-L., Karapanagiotis, I., Magiatis, P., 2013. Structure elucidation
612 and chromatographic identification of anthraquinone components of cochineal (*Dactylopius coccus*)
613 detected in historical objects. Anal. Chim. Acta. 804, 264–272

614 Staunton, J., Weissman, K.J., 2001. Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. 18, 380–
615 416

616 Sugimoto, N., Goda, Y., Suzuki, J., Kuroyanagi, M., Yamada, T., Toshihira, K., Maitani, T., 1998. Structures of
617 minor pigments in cochineal dye. Nat. Med. 52, 135–139Sharma, K.K., Ramani, R., Mishra, Y.D., 1997. An
618 additional list of the host plants of lac insects, *Kerria* spp. (Tachardidae: Homoptera. *J.* Non-Timber Prod.
619 4, 151–155

620 Tang, Y., Lee, T.S., Khosla, C., 2004. Engineered biosynthesis of regioselectively modified aromatic
621 polyketides using bimodular polyketide synthases. PLoS Biol. 2, e31,

622 Tsuchida, T., Koga, R., Horikawa, M., Tsunoda, T., Maoka, T., Matsumoto, S., Simon, J.-C. J.-C., Fukatsu, T.,
623 2010. Symbiotic bacterium modifies aphid body color. Science 330, 1102–1104

624 Vea, I.M., Grimaldi, D.A., Strimmer, K., Lartillot, N., Ronquist, F., 2016. Putting scales into evolutionary time:
625 the divergence of major scale insect lineages (Hemiptera) predates the radiation of modern angiosperm
626 hosts. Sci. Rep. 6, 23487

627 Wouters, J., Verhecken, A., 1987. The chemical nature of flavokermesic acid. Tetrahedron Lett. 28, 1199–
628 1202

629 Wouters, J., Verhecken, A., 1989. The coccid insect dyes: HPLC and computerized diode-array analysis of
630 dyed yarns. Stud. Conserv. 34, 189–200

631
632

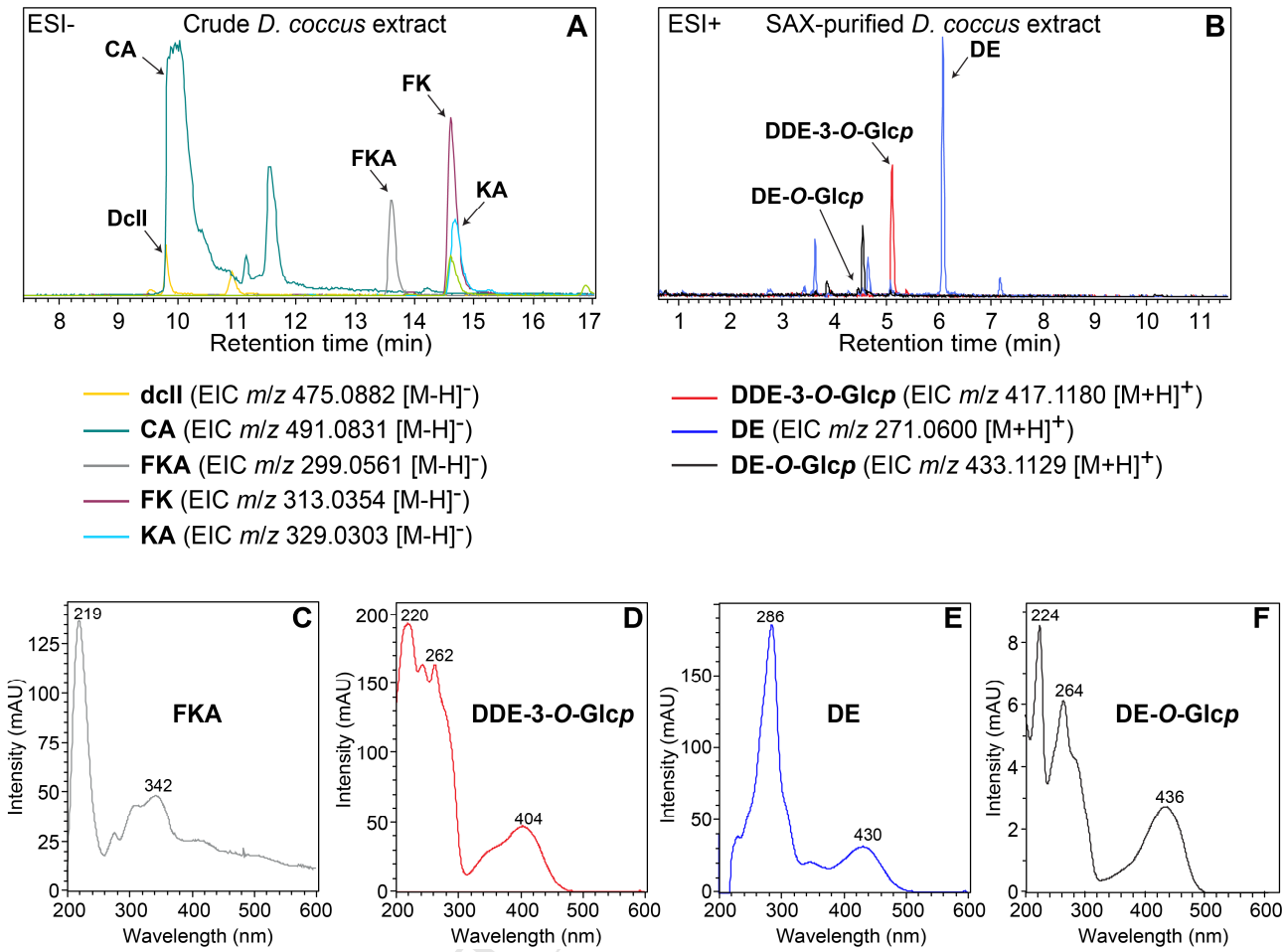
633

634

635 **TABLES AND ARTWORK**

636

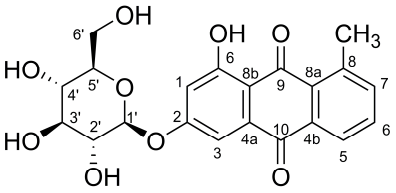
637



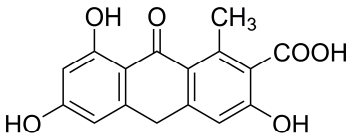
638

639 **Figure 1.** HPLC-HRMS analysis of **CA** related compounds found in *Dactylopius coccus* with annotation of
640 major peaks and UV/VIS spectra for the novel compounds. **A:** Annotated EIC chromatogram for the **CA**
641 related compounds identified in the raw extract of *D. coccus*, coloring scheme for the EIC is shown below
642 the chromatogram. **B:** Annotated EIC chromatogram for **CA** related compounds identified in the SAX
643 purified *D. coccus* extract, **C-F:** UV/Vis spectra for **DE** and the new compounds **DDE-3-*O*-Glc*p***, **DE-*O*-Glc*p***,
644 and **FKA**.

645



5,6-dihydrohydroxyerythrolaccin 3-*O*-β-D-glucopyranoside
DDE-3-O-Glc*p*



Flavokermesic acid anthrone
FKA

646
647
648 **Figure 2.** Structures of the new compounds **DDE-3-*O*-Glc*p***, and **FKA**
649
650

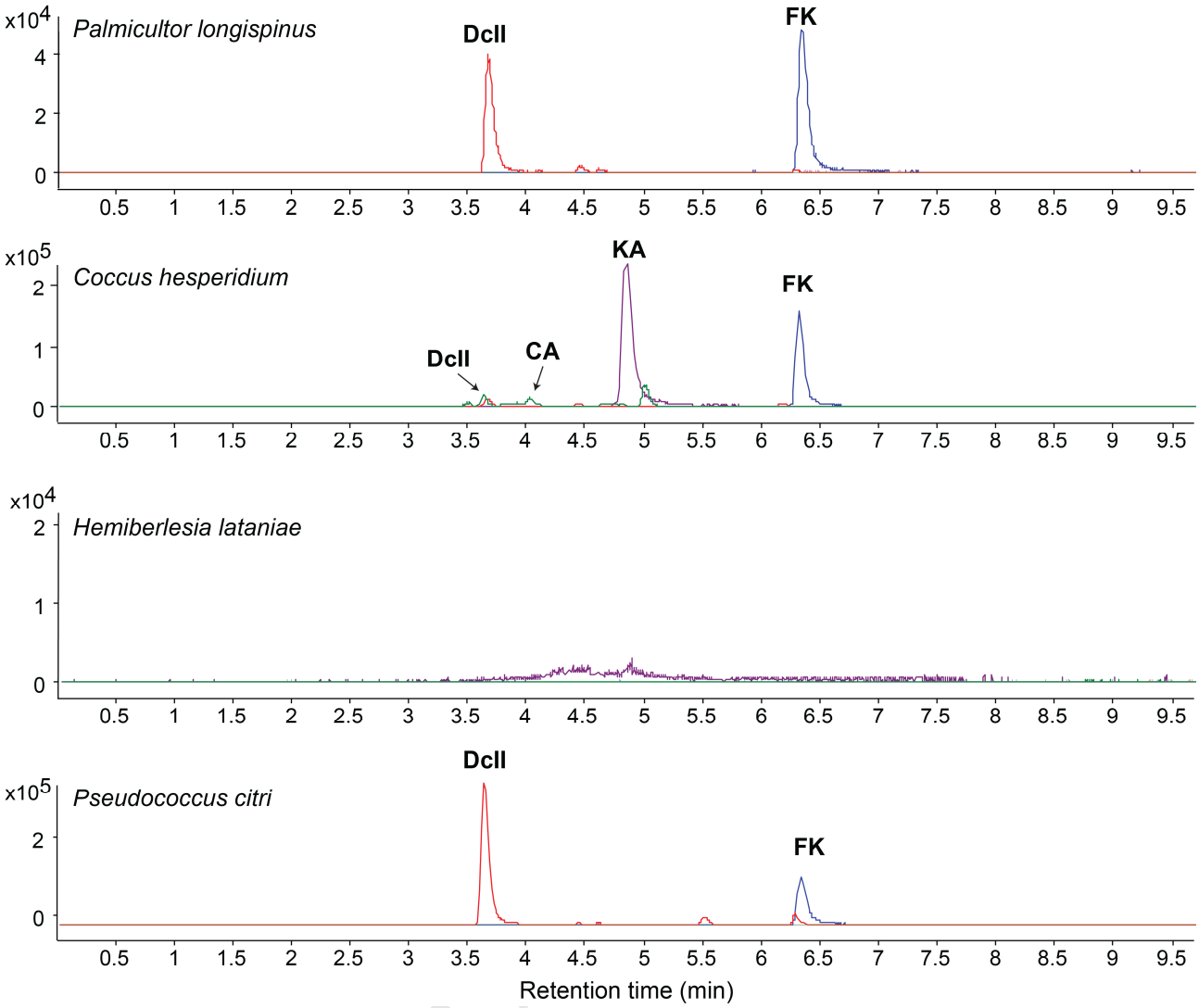
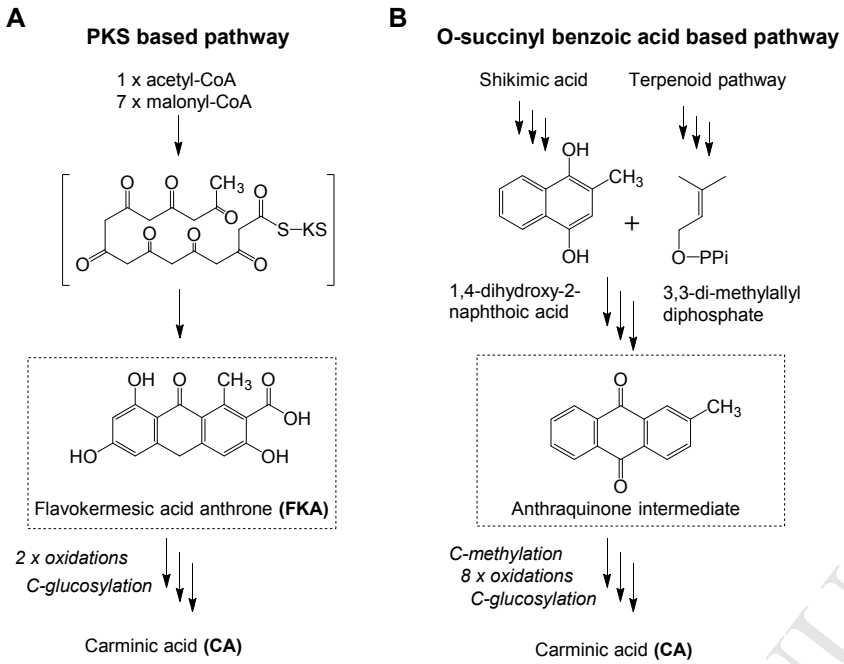


Figure 3. Targeted metabolite analysis of coccid dyes in representatives of the *Coccus*, *Pseudococcus*, *Hemiberlesia*, and *Palmicultor* genera. Extracted ion chromatograms at *m/z* 313.0351 ± 0.01 equal to the $[M-H]^-$ ion of **FK** (green), at 329.0246 ± 0.01 for the $[M-H]^-$ ion of **KA** (purple), at 475.0875 ± 0.01 for the $[M-H]^-$ ion of **dcII** (red), and at 491.0832 ± 0.01 for the $[M-H]^-$ ion of **CA** (green).

657



658
659

660 **Figure 5.** The two theoretical biosynthetic schemes that can lead to the formation of carminic acid with the first stable
661 tricyclic intermediates shown in boxes. Panel A shows the polyketide based pathway while B shows the O-succinyl
662 benzoic bases mechanism.

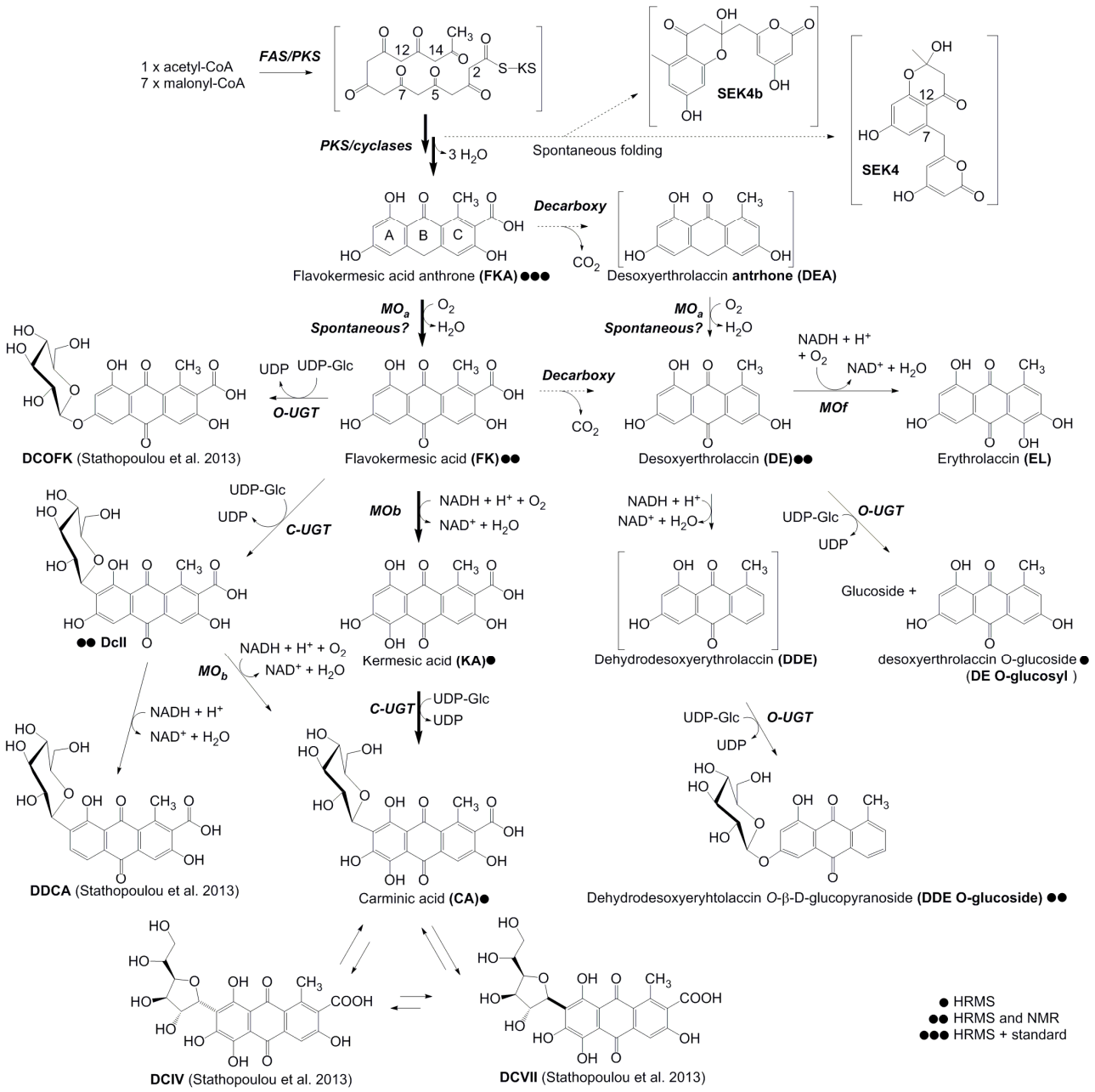


Figure 5. Model for carminic acid biosynthesis in *Dactylopius coccus*, including predicted enzyme types, substrates, and co-factors. Compounds in brackets represent hypothetical intermediates that have not been detected. Signatures: ● compound identified based on HRMS data; ●● compound identified by HRMS and NMR data; ●●● Compound identified based on authentic standard and HRMS data. The numbering of carbon atoms refers to their position in the polyketide backbone, counting from the carbon closest to the enzyme upon completion of the catalytic program. Enzymes: OxidoR = oxidoreductase; MO = monooxygenase (three different a, b and f); UGT = UDP-glucose dependent membrane bound glucosyltransferase; PKS/FAS = polyketide/'mutated fatty acid' synthase; Cyclase/aromatase = small-molecule-foldases as found in bacterial type II iterative PKS systems.

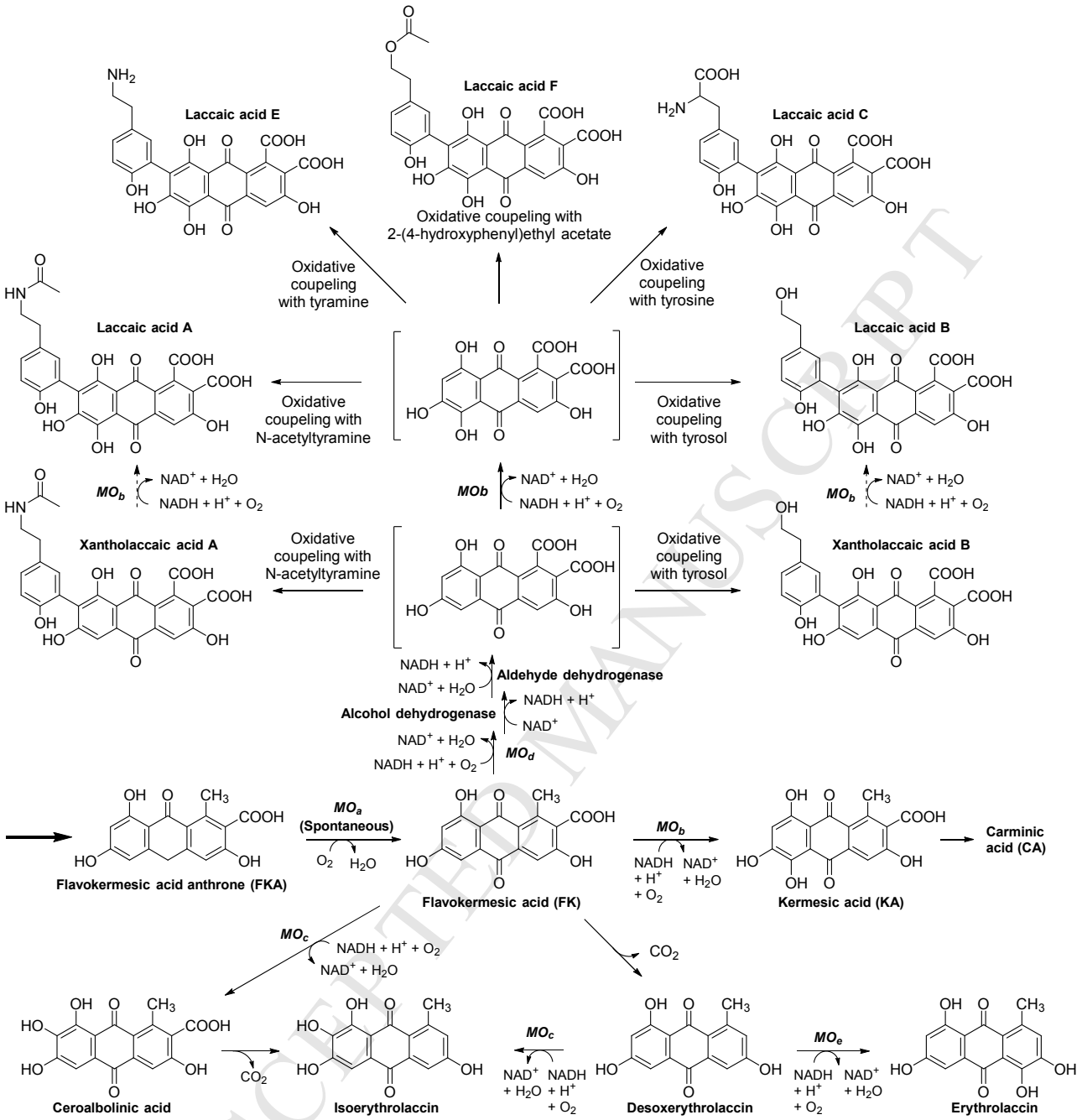


Figure 6. A summative meta-model for the biosynthesis of coccid dyes in scale insects, including predicted enzyme types, substrates, and cofactors. Compounds in brackets represent hypothetical intermediates. The meta-model includes one decarboxylation step and six different monooxygenases (MOa-f) reactions, which are likely catalyzed by different enzymes, based on different positions in the **FK** core that is modified. The oxidative coupling reactions, branching out from the two hypothetical intermediates, are likely unspecific reactions that occur via highly reactive radicals generated by the insect's innate immune system.

- Two novel coccid pigment intermediates isolated from *Dactylopius coccus*, including flavokermesic acid anthrone.
- Detection of the anthrone for the first time provides solid evidence for biosynthesis via the polyketide biosynthetic pathway.
- Coccid pigments is much wider spread within the Coccoidea superfamily than previously believed, suggesting a common genetic basis and evolutionary origin.